AISB Journal 2(1), © SSAISB, 2010
http://www.aisb.org.uk

Non-Verbal Behaviour for Believable Synthetic Agents

John Shearer*, Philip Heslop*, Patrick Olivier* and Marco De Boni†

* Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom, *john.shearer@ncl.ac.uk ; philip.heslop@ncl.ac.uk ; p.l.olivier@ncl.ac.uk*

† Unilever Corporate Research, Unilever R&D Colworth, Sharnbrook, Bedford, MK44 1LQ, United Kingdom, *marco.de-boni@unilever.com*

Abstract

Realism for synthetic characters, both in computer games and conversational agent mediated applications, requires both visual and behavioural fidelity. One significant area of synthetic character behaviour, that has to date received little attention, is non-verbal behaviour. In identifying the scope and participants of non-verbal behaviour in computer games we first review the range of spatial and task scenarios that are relevant. We then select the principal categories of non-verbal behaviour: proxemics, gaze, gesture, self-adaptors and passive communication – summarising their role in communication and propose their incorporation in the design of non-player characters. In conclusion we review the question of how to capture the non-verbal behaviour of players or provide them with interaction techniques that might facilitate non-verbal communication with players and non-player characters alike.

1 Introduction

Perhaps the most significant current challenge for synthetic agents in games is the integration of narrative and intelligent character behaviour with a view to further enhancing user engagement. The development of non-player characters has primarily focussed on high-level coordination of non-player character physical behaviours (typically through the characterisation of state space) and little attention has been paid to low level interactions between characters (both non-player characters and player characters) such as different forms of verbal and non-verbal behaviour.

Realistic interaction between characters is a challenging problem. In simple terms it requires non-player characters to have sophistication in their beliefs, desires, and intentions equivalent to the expectations that a player might have of a real character (within the narrow domain of the game). Although, one should keep in mind the possibility of an 'uncanny valley' (Mori, 1970) – that, as characters become more human-like, with a positive emotional response from a human, there becomes a point where the response suddenly becomes strongly negative – the character is almost human, but seems unusual or strange – something is perceived as being not quite right. This is clearly contrary to the desired effect.

Crucially, non-player characters must also have beliefs as to the beliefs, desire and intentions of other characters, and therefore require some ability to monitor the activity of other characters for intentional behaviour. This then allows for communication (the intentional influence of other's beliefs). This requires a non-player character to distinguish between behaviour that is, and is not, intentionally influencing another's beliefs, as many behaviours are interpreted differently depending on the context. For example, opening a door can be an action (enabling continuation of a route) and communication (showing another person respect). Humans (and many higher mammals) are extremely good at separating basic actions from actions intended to communicate (as a conscious act or otherwise), and, in the course of social interactions, people constantly analyse each other's behaviour (e.g. speaking, moving, gesturing, touching) to understand their (the other's) beliefs, desires and intentions. It has been shown that people also do the same for artificial entities (Reeves and Nass, 1996). That is, people have a tendency to attribute intention to all things they interact with, even if those objects have no possible mechanism from which intention might derive (according to our present scientific understanding of the world), for example, shouting at a household object if it is not working correctly, as if the object intended to go wrong. Inevitably, players ascribe intention (often very complex intention) to non-player characters even when the non-player character's behaviour, and/or the underlying control architecture, is very limited.

Non-player characters in present-day games often incorporate a restricted notion of desire and intention, but have no significant ability to perceive the desires and intentions of other characters (and thereby infer their beliefs and other aspects of internal state). Thus, intention is usually exhibited through direct action – in a first person shooter non-player characters display their intention to frag (kill) a character simply by shooting at that character. Indeed, in games where the role of non-player characters is not so clearly defined, players quickly habituate to repetitive behaviour as this "supporting cast" mills around according to simple rules or scripts.

A discussion of the development of interactions in all areas of games is beyond the scope of our analysis. Instead, we address communication between characters, specifically non-verbal communication through gesture, facial expression, eye gaze and other bodily movement. That is, we address non-verbal interactions between characters that are not direct actions (i.e. not actions such as attacking). Most games portray the intentionality and personality of characters with cut-scenes – scenes inserted, pausing game-play, at various points of the game – with fully animated characters using non-verbal and verbal behaviour in an attempt to provide an engaging narrative for a game. In contrast, during actual game-play a character's behaviour is much simpler, and very little intentionality is usually displayed.

The advent of high performance 3D graphics hardware and software, and high quality audio has given rise to the use of geometrically detailed, skeletally animated characters, that exhibit some non-verbal behaviour. For example, *Half-Life 2* (Valve, 2004) makes extensive use of non-verbal behaviour in non-player characters, to the extent that player engagement and narrative development can occur effectively without the use of cut-scenes. Non-player characters attempt to talk with the player and when engaged in talking to a player character they will: move around to maintain eye contact during the conversation; respond to the players actions; and show expressions and gestures appropriate to the game-play.

While the characters within *Half-Life 2* appear to have desires, personality and intentions, their actual behaviour is, in fact, highly scripted. Most non-player characters have only a relatively small set of simple behaviours and a number of complex scripts that occur only once or twice. These complex behaviours replace the traditional cut-scene, but without the need to break game-play. However, the integrity of the non-verbal behaviour relies on the skill of the artists in coordinating the non-player character's action with the game-play and the non-player character's speech. This can clearly be seen when observing non-player characters' interactions with each other. Aside from pre-scripted cut scenes, non-player characters do not interact with each other, other than to avoid collision. Even then, they still walk on paths that lead to collision, but evade each other at the last moment (sometimes accompanied with a sound effect along the lines of 'watch where you are going'). Indeed, following an individual non-player character through the environment reveals that they are predominantly on pre-fixed paths that simply enter the environment at one point and walk off at another. Some of the inhabitants even walk in a constant looping path around the environment.

Similar behaviour is exhibited in other games, such as *Grand Theft Auto 3* (Rockstar, 2002). Although more densely populated, the non-player characters exhibit similarly aimless behaviour, reacting only to certain events (such as explosions or being attacked) but otherwise wandering through the environment with no apparent destination. There are games in which the behaviour of non-player characters is more sophisticated, such as *The Sims* (Maxis, 2000), where the characters fall half-way between being user controlled and AI controlled. The characters converse with each other and utilise objects in their environment. Though, even in this case the characters exhibit only the desire to satiate certain undesirable internal states, such as hunger and tiredness, by using the objects and characters around them. They have no longer term and more complex goals, and the conversations between characters do not actually convey any specific information but rather the act of conversation itself satisfies some current short-term need; the characters can obtain the same effect by using an object in the environment, such as reading a book or listening to the radio. In practice the scope of application of non-verbal behaviour extends far beyond the replacement of the cut-scene and in the next section we consider the applicability to games of the relevant spatial and task contexts over which non-verbal behaviour is known to be utilised in humans.

2 Spatial and Task Context

Communication can be considered to occur in four different task contexts: cooperation, coaction, competition and conversation (Knapp and Daly, 2002). In other words, communication occurs in order for some number of parties to: perform a task together (cooperation), to exist in the same vicinity (coaction), to perform a task at the expense of another (competition), and to entertain or pass on information (conversation). Communication varies across these different contexts and also with the physical proximity of the communicating parties. As already described, non-verbal behaviour provides information as to the beliefs, desires, and intentions of a character, or alternatively it can be considered as providing indicators as to another character's cognitive, emotional, physical, intentional, attentional, perceptual, interactional and social status. A set of distinct but common communication situations (spatial and task contexts) for computer games is illustrated in Figures 1-6 using screenshots from *Half-Life 2*.

Figure 7 maps out the range of spatial and task context for these examples. Synthesizing non-verbal behaviour for conversational partners at close physical proximity is particularly difficult, due to the full movement of a character (both body and face) being visible in detail to the player. Furthermore, people are highly attuned to interactions in intimate, personal, and social spaces and are sensitive to many subtle cues and nuances in non-verbal behaviour. Thus, at close proximity players are very aware of errors, unrealistic, or unnatural behaviour in non-player character. At further distances less detail of a character's movement is apparent. Moreover, there is a significant transition in non-verbal behaviour from situations where intimate verbal communication is possible to those where it is not. The sensitivity of non-verbal behaviour to proximity is due to a number of factors, including the more public nature of non-verbal gesture in open spaces, and the requirement on particular physical behaviour to carry the full communicational load (e.g. subtleties in gaze and facial expression are not visible at a distance).

Figure 1 shows an example of cooperation in intimate space. The male character demonstrates his attentional state – that he is attending to the female character – with his body orientation, face orientation, and gaze direction. Of course, people are rarely static, but different non-verbal channels (e.g. face orientation, body orientation, gaze direction, body position) are closely coordinated in demonstrating attention. Thus, the male character could look away but still communicate his attention sufficiently through his posture and pose. In an interaction between unfamiliar participants, however, strong or constant facing or looking at a person is widely considered an aggressive signal. It is considered rude, or at least off-putting (Knapp and Daly, 2002). Figure 1 also illustrates non-verbal behaviour using facial expressions and kinaesthetic (touching) behaviour.

Figure 5 illustrates a situation at the other end of the spatial scale, cooperation at a distance between the player and a non-player character (in fact, navigation and negotiation, a subset of cooperation). The non-player character shown and the player will collide if they do not arrive at an agreement as to how to pass one another and communicate this – the characters must cooperate through the use of non-verbal behaviour to resolve a potential conflict. In the real world, people in this situation use a range of subtle non-verbal mechanisms such as gaze and body turning to initiate and mutually negotiate space. Non-player characters in *Half-Life 2* will avoid the player, but will not exhibit non-verbal behaviour in doing so and simply move around the players as they approach. Without non-verbal behaviour it is difficult for the player to decide which way to move out of the way (indeed they do not need to) and this absence of social convention (and the ability to break them, to invite conflict) that both undermines the engagement of players with the game and limits their expressivity.

Between the proximal and distant spatial scales are social spaces and figure 6 is an example of conversation in a social space. Here non-verbal behaviour facilitates a number of aspects of the interaction (and the dialog in particular) including the mediation of conversation flow, such as whose turn it is to speak (interactional state). Turn-taking mediation is a complex coordination of behaviours, but in simple terms speakers provide opportunities to allow the listener to take a turn (such as, a slightly prolonged pause, or a look up into the eyes), at which point other listeners can, if they choose, take a turn. If not, then the speaker will continue. Additionally, the 'speaker' can indicate that they would like to speak, with signals such as increased eye contact, leaning forward or standing taller (Duncan and Fiske, 1977). Turn-taking mediation is not required in *Half-Life 2* because the game developers have not allowed the player to speak, but it is potentially a very important component of games that hope to include natural language interactions (particularly spoken interaction) between player and non-player characters.

Finally, figures 2, 3 and 4 illustrate the remaining task contexts: coaction, conversation and cooperation, and competition. Characters sharing the same approximate area of space engage in coaction behaviour, corresponding to mutual monitoring – this can be interpreted as communication by virtue of the fact that watching a character implies that you might react to it – that is, there is an implied reason (intention) for watching. Coaction can be considered the default task context, which develops into the other contexts. Competition contexts give rise to distinctly different forms of non-verbal behaviour from other contexts, but these still serve to communicate internal states. In figure 4 the raised baton serves to communicate"you have crossed a line – back off or I will hit you".

Figure 1: Cooperation in intimate space

Figure 2: Coaction in social space

Figure 3: Conversation in personal space

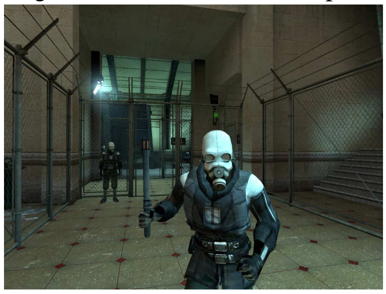

Figure 4: Competition in personal space

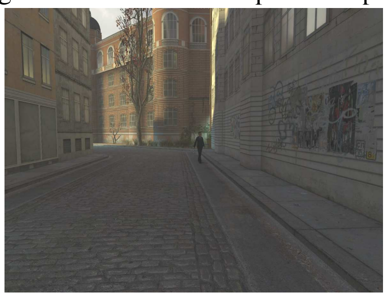

Figure 5: Cooperation at a distance

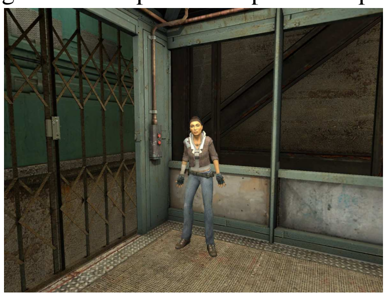

Figure 6: Conversation in social space

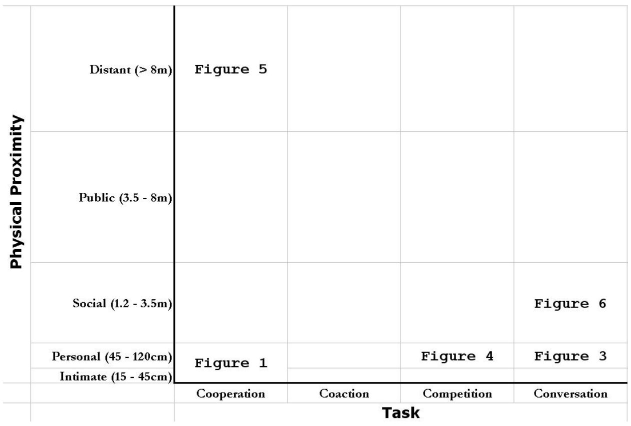

Figure 7: Task and spatial contexts.

Table 1:

3 Non-Verbal Behaviour

Non-verbal behaviour in synthetic characters should be based on the equivalent behaviour exhibited by real people in human-human interactions. Hence, the forms of non-verbal behaviour that are particularly relevant to the synthesis of non-verbal behaviour in virtual characters include:

- Proxemics – (the use and arrangement of the self in the physical world).
- Gaze – (where the eyes are looking and pupil dilation).
- Gesture – (movement of hands and head).
- Self-Adaptors – (movements that serve to alter the self).
- Passive Communication – (communication that occurs without a specific action).

It should be noted that there are very significant individual differences in the precise nature of their non-verbal behaviour. Non-verbal behaviour is influenced by age, gender, social status/hierarchy position, culture, presence of others, illness and physical ability as well as the spatial and task context. This presents problems for both the study and the synthesis of non-verbal behaviour, but also furnishes a potentially rich behavioural vocabulary by which the individuality of characters can be developed. Fidelity in non-verbal behaviour requires that each distinct non-player character employs its own characteristic behaviour in this respect. In the following sections we characterise non-verbal behaviour in people, generally from a Western cultural perspective, with a view to identifying which aspects should be addressed in the design of engaging virtual characters for computer games.

3.1 Proxemic Behaviour (Proxemics)

Proxemic behaviour is defined as the closeness and arrangement of the self in physical world compared to others (Hall, 1966). This is the use of personal space and territory, and although spatial measurements are sometimes difficult to judge in game environments (people rarely feel their personal space is invaded in games), the use of space does have an impact on player behaviour. Moreover, as games become more immersive (due to improved graphics, larger screens, head-mounted displays, immersive environments, and surround sound) and spoken dialog with non-player characters becomes a possible occurrence, the impact of proxemics will increase. As discussed, non-verbal behaviour varies significantly with distance (and spatial arrangement), so its consideration is therefore important in the synthesis of non-player character behaviour.

Proxemics also exhibit some of the largest (but consistent) cultural variations, especially in conversation. In addition to distant space for interaction at over 8 metres (25 feet), where people still interact but do not conduct conversations, we can define, in order of increasing distance (or decreasing proximity), the spaces for UK/US culture (Hall, 1966):

- intimate space for embracing, touching or whispering (15-45 cm, 6-18 inches);
- personal space for interactions among good friends (45-120 cm, 1.5-4 feet);
- social space for interactions among acquaintances (1.2-3.5 metres, 4-12 feet);

- public space used for public speaking (over 3.5 metres, 12 feet).

Proxemic behaviour stems from the idea of territory. Territories vary dynamically and are dependent upon many different factors, but the categories above are likely to be sufficient for basic synthesis. Synthetic characters should use proxemics in the same way that people do – in order to be socially correct use: social space for interacting with relatively unknown people, personal space for people known well and intimate space with people known very well. Violation of these rules is socially incorrect. Invasion of personal space can be intimidating, flirtatious, or can be the result of interaction between participants from different cultural backgrounds. Whilst the requirements of a task, such as needing to be close to attend to a wound, may override the usual norms, in such circumstances the progression through the spaces is still mediated by non-verbal behaviours.

The synthesis of complex proxemic behaviour requires further parametrisation with respect to types of territory: primary, secondary and public (Altman, 1975). Primary territories are the exclusive domain of the owner (such as a home), secondary territories are those felt to be partly owned (such as the local pub), and public territories those available to almost anyone for temporary ownership (such as a park bench). Temporary ownership means a person behaves very differently towards that object or space while they 'own' it (Knapp and Daly, 2002), it becomes part of their territory.

3.2 Gaze and Eye Based Communication

The use of the eyes is an important component of human-human communication, and it involves far more than just what a person is looking at. Peoples' eyes are constantly moving from place to place (not smoothly, but jumping from one locus of attention to another). As Knapp and Daly (2002) note, "we associate various eye movements with a wide range of human expressions: downward glances are associated with modesty; wide eyes with frankness, wonder, naivete, or terror". Where the eyes are looking (or gaze) is the primary (and most obvious) form of communication by gaze, but people are also sensitive to, and react to, pupil dilation/constriction. There is also a close relationship between gaze behaviour and facial expression (not considered here).

Kendon (1967) identifies four functions of gaze behaviour (in addition to looking at specific items for information gathering), and Knapp and Daly (2002) built on this, classifying five functions of gaze as:

1. Regulating the flow of communication.
2. Monitoring feedback.
3. Reflecting cognitive activity.
4. Expressing emotions.
5. Communicating the nature of an interpersonal relationship [added by Knapp].

Within the context of a non-player character, the five categories pose challenges for synthesis that are considerably more challenging than simply having the character look at where it is interested. The regulation of communication flow, gazing briefly at another person (specifically at the face) establishes an obligation to interact; further and longer gazing shows a desire to increase the level of interaction; while decreased and shorter gazing desires a decrease in the level of interaction. During an interaction eye glances serve as turn-taking signals and also highlight grammatical breaks, conceptual unit breaks, and the ends of utterances (a sequence of speech separated from another by a marked gap). These glances also allow feedback on the interaction by monitoring the reactions of the other person.

When under increased cognitive load (for example trying to process difficult or complex ideas) both listeners and speakers tend to look away. It is thought that this averted gaze reflects a shift in attention from the external to the internal. Additionally, there is evidence that the eye gaze direction under this condition changes with different forms of cognitive load, according to the active hemisphere of the brain (Ehrlichman and Weinberger, 1978; Weisz and Adam, 1993; Wilbur and Roberts-Wilbur, 1985).

The eyes are also a site for the display of emotions: surprise; fear; disgust; anger; happiness; sadness, in addition to blends of these and more complex emotions. Interestingly, in certain contexts emotions displayed with the eyes will not always match the facial emotional expressions (e.g. during emotional masking) and can be very transitory. However, people are adept at detecting emotional state from the eyes, and there is evidence that different emotions are in fact detected from different areas around the face (Ekman et al., 1971; Ekman and Friesen, 1975), but that it is the facial area around the eyes that displays the emotion, not the eyes themselves.

Finally, eye gaze can communicate the nature of an interpersonal relationship. Gazing and mutual gazing is found most in conversations with a high-status addressee, lower with a very high status addressee, and minimal with a very low-status addressee (Hearn, 1957; Efran, 1968). Aggressive encounters and interactions between lovers or mothers and babies have extended periods of mutual gazing.

3.3 Gesture

Gesture is body movement that serves to communicate, mainly involving movement of the hands and the head. There are many forms of gesture, but of particular relevance for non-verbal behaviour synthesis in non-player characters are emblematic gesture (gesture with specific meanings that occur without speech) and spontaneous gesture (hand and head movements that occur with speech). Emblematic gestures are well defined in both their form and meaning and are therefore readily synthesised with standard skeletal animation and scheduling frameworks. For example, the 'come here' gesture is performed using the moving of a finger, fingers, hand, hands, arm, or arms towards the body from the direction of the addressee (often in a repetitive form). Synthesis of spontaneous gesture for non-player characters is significantly more challenging.

Spontaneous gesture is performed by people while speaking, in synchrony with the speech and is generally made with the head or hands. When needs be people are apt to use any available body part, or even the whole body (e.g. pointing with a foot when one's hands are full). Spontaneous gesture continues to be studied in depth by the psychology and psycholinguistics communities, and though most studies are descriptive in nature, recent research on growth points is leading to theories of how gesture relates to other cognitive processes (McNeill, 2005).

It has been found that the gesture stroke (the semantic, or meaningful part) of a gesture commonly coincides with the peak phonological stress – the most emphasised phoneme – of the speech stream. Gesture is tied closely to the underlying speech and both speech and gesture are widely believed to be generated from a single underlying conceptual representation (McNeill, 1992). Indeed, spontaneous gesture can be complementary, supplementary, or contrastive to the speech. In other words, gesture can re-iterate or emphasise the speech, add information to the speech, or communicate something contradictory (or slightly different) from the associated speech. In contrast to speech, gesture has few constraints on how it is constructed. As McNeill (1992) notes "the important thing about gestures is that they are not fixed. They are free and reveal the idiosyncratic imagery of thought." McNeill (1992) identifies five categories of spontaneous gesture: Emphatic, Deictic, Cohesive, Metaphoric and Iconic.

Emphatic gesture (also known as beat gesture, or baton gesture) provides emphasis to parts of speech: phrases, words and phonemes. Emphatic gesture consists of just two movement phases (up/down, in/out, left/right) with the transition from one phase to the other (such as, up to down) being the point of emphasis. These have little variation in form other than the scale and speed of the phase transition, with larger, faster transitions indicating more emphasis (within an individual). Emphatic gesture can, and frequently does, utilise all body parts, especially the head and hands, but additional movement of more of the body provides further emphasis. This form of spontaneous gesture is distinct from other forms in that it can overlay any other gesture as it indexes a part of speech rather than providing semantic content (though it is also frequently used independently).

Deictic gestures are simply pointing actions that refer to an object or objects, generally using fingers, hands, or head. This is complicated by the fact that deictic gesture can, in addition to referencing concrete objects, reference more abstract objects, such as where an object was previously, the physical space referred to previously with the introduction of an idea or object, or almost any abstract space, including time. For example, when describing a cartoon involving two characters people will often reference one specific area of their gesture space for one character and another separate area for the other (McNeill, 1992). It should be noted that deictic gestures can take the form of whole body movement towards a space as well as the orientation of body parts.

Cohesive gestures serve to connect related parts of discourse that are temporally separated. For example, when listing items people often provide an emphatic gesture on each item. The emphatic gesture marks each item, while the repetition of the same gesture form connects them together to say 'here's one, and another, ...'. Cohesive gesture usually consists of the repetition of a specific gesture and so they require the use of other gesture forms (which could simply be emphatic gesture).

Metaphoric and iconic gestures are in essence animated representations. An iconic gesture is a pictorial animation of a concrete entity or action, while metaphoric gesture animates an abstract concept as if it were a physical object. For example, a speaker saying 'and he bends it way back' while illustrating the action of bending a stick is performing an iconic gesture (McNeill, 1992). In contrast metaphoric gesture occurs in situations such as when a speaker says 'I had this great idea' and marks the 'great idea' with a cupped hand gesture (i.e. metaphoric container). Metaphoric gesture takes abstract ideas and grounds them in real entities. In practice only a few types of object are portrayed, mostly containing or enclosed objects, though it is often not clear what object is being portrayed.

Iconic gestures are probably the most sophisticated and developed class of gesture; each gesture attempts to portray some aspect(s) of a situation, event, or object in the physical world and therefore absorbs much of the complexity of the physical world. Typically the gesture portrays the most important and semantically salient features in the context of the interaction, and therefore a gesture about a specific object, such as a teacup, can vary significantly depending on the context. For example, a person talking about a teacup in context of drinking tea may perform a gesture of lifting a teacup by its handle and 'drink', while if the context was about how much tea was in the cup, the gesture would be distinctly different – perhaps portraying the size of the teacup or the level of the tea as opposed to how it is lifted. Iconic gesture is the form of gesture that mostly allows for the addition of extra information in the gestural channel and almost anything is permitted.

For objects or ideas that are common in interaction, the gesture forms portraying them can often become stylised, and eventually can become symbols, akin to emblems.

In addition to gesture accompanying speech, gesture can be used in a similar way to eye gaze and many vocal signals to regulate communication flow and the rhythm of interaction (Knapp and Daly, 2002). Head nods are the most frequent form of these gestures, but hand and body gestures can also serve for flow regulation. Unsurprisingly, flow regulation gesture frequently coincides with flow regulation signals in other channels.

3.4 Self-Adaptors

Self-adaptors are movements that serve to change the self, such as scratching an arm; as such they are typically not intended to serve a communicative function – their purpose is to adjust something about the self. Certain reading as to the internal state of a speaker, based on their use of self-adapters, have been proposed for communication. Self-adaptors can take on many forms and are distinctly idiosyncratic. The meanings assigned to them by conversational participants and bystanders vary considerably. For example, flicking hair out of the face, while a practical movement, is often interpreted as a flirtatious behaviour.

3.5 Passive Communication

Finally, another consideration to be taken into account is passive communication – communication that occurs without a specific action on the part of the communicator, and usually takes place at the co-action level of communication, and at a greater distance than conversational communication. The communicator expresses information about themselves simply by their physical appearance, demeanour and observed behaviour. These factors can be considered by the user (or indeed by other non-player characters) before initiating a more direct method of communication. For example, a large aggressive character might be approached with more caution than a small timid looking character. Similarly, witnessing a character commit some violent act would instil more caution in the user than a character who has been doing something less impactful. Passive communication is often unintentional at the time of communication itself, but can be premeditated, for example wearing a certain type of clothing may convey a specific message.

4 Modelling Synthetic Agents

4.1 Behavioural Models

The objective of behavioural models is to isolate a particular behaviour or group of behaviours of the subject and create a simple mechanism to specifically simulate behaviour or behaviours. The goal is not to produce a full representation of the subject, restricting the simulation to the specific actions the chosen behaviours require.

Reynolds (1987) paper neatly outlines this philosophy and how such a system might be implemented. Reynold's aim was to simulate the flocking behaviour of birds (or the herding of animals or shoaling of fish which are visually similar behaviours). In order to accomplish this he took the approach of defining the behaviour itself as a series of rules that each participating 'boid' would follow. He did not for example attempt to simulate the sensory system of a bird or fish, or attempt to create a functioning 'brain' that reasons deliberatively.

Reynolds took the notion of a particle system (Reeves, 1983) and its physical rules (for example representing gravity) and augmented them with 'social rules', such as 'try to fly in the same direction as near neighbours' or 'try to fly towards the centre of the flock'.

Following the rule-based behaviour, the boids did eventually exhibit realistic behaviour, but it was not simply a result of defining a set of rules and allowing the boids to follow them. The interaction of each rule was a difficult factor to consider, especially if rules contradicted each other. Taking the average of rules was not enough, Reynold's gives the example: "Consider flying over a grid-work of city streets between the skyscrapers; while 'fly north' or 'fly east' might be good ideas, it would be a bad idea to combine them as 'fly north-east'." Instead, a rule hierarchy was created and weights attached to each rule. Each rule that was triggered had its weight added to a rule accumulator, and its effect (on acceleration) was added to another accumulator. If either accumulator reached a maximum value, the rest of the rules were ignored for that pass.

Behavioural modelling can produce a simple and elegant solution to portray a specific phenomenon, and its simplicity allows many synthetic characters to be simulated at once. However as it is focused on a specific behaviour, the synthetic characters are restricted, and do not fully encapsulate all the behaviours that may be expected from a rich synthetic character.

4.2 Cognitive Models

While behavioural modelling can produce realistic, if restricted, behaviour, cognitive modelling takes a more holistic approach to simulating a synthetic character. Funge et al. (1999) characterizes that a cognitive model is "What a synthetic character knows, how information is acquired and how the synthetic character uses it" – a model of memory (knowledge), senses (acquiring information) and ultimately the decision making process (how to use the information).

Funge describes a 'cognitive layer' controlling the behaviour of a synthetic character (and ultimately behaviour goes on to control physical which in turn determines kinetic forces which finally specifies geometric changes). The cognitive layer would make choices and decisions based on its memory and sensory input to choose which behaviour to exhibit. The modelling of such a process readily incorporates learning. The cognitive model can be trained through experience to make decisions that are not initially programmed into the character. This allows the modelling of expandable and adaptive synthetic characters that potentially have a rich set of behaviours that cover many situations. The computational cost of reasoning with such cognitive models is substantial. If there are many sensory inputs that need to be considered, or a large knowledge base to process, a character may require a complex and time consuming deliberative process. This overhead can limit the number of concurrent characters acting in a virtual environment.

When specifically considering modelling a human-like synthetic agent, the concept of a cognitive model can be extended beyond a deliberating 'brain' to incorporate models for human senses, memory and other human cognitive functions (Hartmann et al., 2002; Peters et al., 2006). The purpose of adding these elements is to recreate realistic socially enabled synthetic characters by emulating the internal processes that occur in humans. The agents produced are specifically embodied, multi-modal (in that they incorporate visual gesturing as well as verbal communication) conversational agents. They have sophisticated sensory inputs, for example vision is based on processing of actual renders of the virtual environment from the viewpoint of the character, with image processing techniques used to obtain areas that require the characters attention (Peters and O'Sullivan, 2002). Synthetic agents use this technique to perceive other characters' gaze during communication, pick up gestures and even read facial expressions.

Sensory, long term and short term memory are also simulated. The layering of memory allows data to be prioritised (by having it in short term memory), allows repeated data to not be re-stored (by referring to long term memory) and allow sensory data to be attended to (if in sensory memory) or ignored (if short term memory is full).

The synthetic agents also have a complex decision making process, or 'theory of mind'. This is where sensory and memory data is evaluated to determine what to attend to and which behaviours to exhibit. The communication between characters is handled by 'reading' visual and vocal cues and transmitting the same when indicating turn taking etc. The decision making process uses the sensory data to pick up cues and decide what to do next. To introduce individualism into the synthetic characters, reactions and animations are varied, which leads to different sensory data in the communication flow and hence different resulting reactions. The result is richer more interactive conversational synthetic characters that talk, gesture, gaze and respond realistically during conversation. However the processing overhead means that the system is currently limited to small groups of characters, and conversations with more than two characters look awkward and lethargic.

4.3 Conversational Non-Verbal Behaviour Synthesis

Conversational non-verbal behaviour in synthetic characters can be generated by analysis of the data stream being communicated (usually text based). Cassell et al. (2001) introduced a system that allowed animators to generate realistic non-verbal behaviours to accompany verbal communication by performing linguistic and contextual analysis of the text to be vocalised. The Input text is separated and tagged by the system, so that different animation fragments can be played to represent the non-verbal behaviour associated with pertinent sections of the text. As well as creating textual input, animators are required to create scripts that define the set of eye, face, head and hand behaviours and associate them with phrases.

Olivier et al. (2006) implements a similar system which utilises natural language processing to analyse text typed by a user conversing with an Embodied Conversational Agent (ECA). This requires the synthetic character to perform the non-verbal behaviours in real-time and in synchrony with the utterances.

4.4 Crowd Modelling

When modelling the behaviour of larger groups of synthetic characters such as crowds, a wide range of methods are used. However most have their basis in either behavioural modelling, with the characters obeying a set of rules, or cognitive modelling where each character is individually modelled and crowd behaviour is an emergent property, or more usually a hybrid between the two systems, where some group behaviour is governed by rules, but individuals have a decision making process as well. Often a behavioural model is seen as a 'top down' approach, where overall crowd behaviour is designed from the outset, and a cognitive model is seen as a 'bottom up' approach, where group behaviour is emergent from the individuals behaviour.

Rymill and Dodgson (2005) take a bottom up approach, where individual characters are modelled based on theories from psychology. Observed human behaviour is used to control collisions and character flow through an environment, incorporating collision avoidance (including avoiding oncoming collisions, overtaking and avoiding glancing collisions while travelling in similar directions). Collision avoidance techniques that do not require a character to change their path, such as the 'step and slide' observed in real human behaviour (Wolff, 1973) are also incorporated. Although individuals were modelled without an overall behavioural model, the observed results were very similar to those found in real crowd movements. For example individuals formed lanes through bottlenecks, and after collision avoidance, characters tend to return to original path rather than creating a whole new path similarly to how real humans behaved in the studies conducted by Daamen and Hoogendoorn (2003); Hoogendoorn and Daamen (2005).

Shao and Terzopoulos (2006) also take an individual approach to modelling crowds. Synthetic characters use a simple cognitive model: they have a series of attributes that are satisfied by different actions in the environment, for example 'hunger' might be satiated by finding a food source in the environment, at which point a preset animation would take place to represent eating, and the hunger attribute would be reset. These attributes decay over time and so need to be topped up by locating the appropriate object or individual, and having different decay rates allowed individuals to behave slightly differently. Sensory inputs allow collision avoidance with objects and other individuals. The model was used to populate recreated archaeological sites with realistically behaving crowds.

Musse and Thalmann (2001) created a control hierarchy for modelling crowd behaviour, a hybrid of behavioural and cognitive modelling. The hierarchy incorporated an overall crowd control, smaller subgroups and individuals. The overall group control was scripted depending on the situation, (examples include evacuation or reaction to user stimulus), but individuals had control over their own actions within the scripted event. Certain individuals were designated as leaders of subgroups and their actions would influence the overall goals and actions of other members of the group. This method produced good performance, with large sizes of crowds being possible, and the behaviour was suitably adaptable that it produced a realistic reaction to user stimulus.

5 Facilitating Interaction

The discussion of non-verbal behaviour above concentrates on the different forms of non-verbal behaviour in an attempt to establish requirements for next generation virtual characters. Mechanisms for the synthesis of such behaviours are inevitably beyond the scope of this analysis but there have been a number of recent attempts to synthesise spontaneous gesture (Kopp and Wachsmuth, 2004), (Olivier, 2004), (Cassell et al., 2001) and gaze behaviour. However, whilst we can imagine a situation whereby non-player characters have sophisticated cognitive models and the ability to both synthesise and interpret non-verbal behaviour, there is no mechanism for the player to communicate non-verbally with either the non-player characters or other players.

Present-day technology allows collection of full data on all aspects of human physical behaviour that may be a channel for non-verbal behaviour. This includes body position, body movement, hand shape, eye gaze direction, pupil dilation, facial expression, vocal behaviour, voice, and a variety of other biometrics. Unfortunately, such data can only be collected accurately using specialised (expensive) invasive equipment. With that commercial consideration, in addition to joystick (or controller), keyboard and mouse inputs it is only reasonable to expect relatively basic additional input devices for computer games in the near future. These devices include webcams and microphones, but also less standard input methods, such as dance mats, light guns, and low point motion capture devices, such as the Gametrack 3D motion tracker.

Within current commercial and technological constraints, three alternatives for the control of player character non-verbal behaviour generation can be identified:

1. Simulation – non-verbal behaviour of human players is simulated as for non-player characters, and is independent of the actual non-verbal behaviour of the human player.

2. Augmentation – as for simulation, but specific controls are given to allow a player to explicitly alter the simulation, such as a slider to indicate how happy the player is, or a button to increase the level of interest in an interaction.

3. Tracking and mapping – the human player is tracked using equipment such as webcams and the coarse features that can be identified are either mapped to explicit controls (as in the case of augmentation) or directly to the character animation.

The challenge for augmentation is to design an interface that is intuitive, non-obtrusive, and useful all at the same time. While the challenges in player non-verbal behaviour data collection are daunting, the challenges of understanding or recognising behaviours or meaning from that data are even more so. Gesture recognition, for example, is in its infancy and mostly addresses the use of gesture as an explicit interaction technique and little research has been conducted into the automated recognition or understanding of spontaneous gesture.

6 Conclusion

Synthetic characters in virtual environments, such as computer games, are currently under utilised as a vechicle for providing believable and immersive experiences. Human-like characters require behavioural fidelity across all task and behavioural contexts, and not only at key points in competition and narrative progression. Behavioural fidelity demands that synthetic characters express themselves to players and non-player characters alike, utilising aspects of non-verbal communication identified in the social psychology literature. Synthetic characters should use appropriate gesture behaviours to augment their verbal communication. Non-verbal behaviour helps aid the flow of conversation, highlights key points during communication and complements and supplements verbal communication through iconic and metaphoric gestures. Gaze is also an important non-verbal communication behaviour that synthetic characters should emulate. Like gesture, gaze regulates flow during communication, also indicating the state of the communicator, reflecting cognitive activity, expressing emotions or even showing the nature of a relationship between two individuals. At greater distances, gesture is also used to communicate intent.

Synthetic characters should behave appropriately depending on their task context. As well as an augmentation to conversation, non-verbal communication occurs between characters when they are competing against each other, when they are cooperating to accomplish some task and when coacting in the same shared social space. They should also display human like proxemic behaviour, taking into account their closeness, and arrangement in regard to, other characters in the shared space. There are physical limitations to communication over different distances or in certain environments, conversations are difficult at large distances, or in a loud place for example. As well as obeying the physical rules for communication, synthetic characters should behave in a socially correct manner – the interaction distance between characters depends on their familiarity and amiability. Proxemic Behaviour is based on the idea of territory, that is ownership of space. Synthetic characters should behave appropriately towards primary, secondary and public territories. The dynamic nature of public territories in particular require characters to obey different social rules depending on if the 'ownership' of space.

Modelling synthetic characters has been approached using several techniques, from behavioural rule-based systems for controlling large numbers of characters, to simulations of aspects of cognition, such as perception and memory. Models have not yet concentrated on the simulation of human-like communication. Modelling the verbal and non-verbal communication behaviours has its own complexities, and in particular requires the ability to perceive and reason about character-character interactions as a third party. Interacting with synthetic characters should also embody the notion of communication through verbal and non-verbal behaviours. Technology is sufficient to allow human users to interact with sufficiently enabled synthetic characters – gesture, spatial position, gaze direction and a variety of other biometrics can all be utilised a inputs to existing interaction technologies – requiring that the synthetic character would be able to recognise these interactions as non-verbal signals.

References

Altman, I. (1975). *The environment and social behaviour*. Brooks/Cole, Monterey, CA.

Cassell, J., Vilhjálmsson, H. H., and Bickmore, T. (2001). Beat: the behavior expression animation toolkit. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 477–486, New York, NY, USA. ACM Press.

Daamen, W. and Hoogendoorn, S. P. (2003). Experimental research on pedestrian walking behavior. *Transportation Research Board annual meeting*, pages 1–16.

Duncan, S. and Fiske, D. (1977). *Face-to-face interaction: research, methods, and theory*. L. Erlbaum Associates, Hillsdale, N.J.

Efran, J. (1968). Looking for approval: Effects on visual behaviour of approbation from person differing in importance. *Journal of Personality and Social Psychology*, 10:21–25.

Ehrlichman, H. and Weinberger, A. (1978). Lateral eye movements and hemispheric asymmetry: A critical review. *Psychological Bulletin*, 85:1080–1101.

Ekman, P. and Friesen, W. (1975). *Unmasking the face*. Prentice-Hall, Englewood Cliffs, NJ.

Ekman, P., Friesen, W., and Tomkins, S. (1971). Facial affect scoring technique: A first validity study. *Semiotica*, 3:337–58.

Funge, J., Tu, X., and Terzopoulos, D. (1999). Cognitive modeling: knowledge, reasoning and planning for intelligent characters. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 29–38, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Hall, E. (1966). *The hidden dimension*. Doubleday, Garden City, N.Y.

Hartmann, B., Mancini, M., and Pelachaud, C. (2002). Greta: Formational parameters and adaptive prototype instantiation for mpeg-4 compliant gesture synthesis. *Proceedings of Computer Animation, IEEE Computer Society*, pages 111–119.

Hearn, G. (1957). Leadership and the spatial factor in small groups. *Journal of Abnormal and Social Psychology*, 54:269–72.

Hoogendoorn, S. P. and Daamen, W. (2005). Pedestrian behaviour at bottlenecks. *Transportation Science*, 39(2):147 159.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:2622–63.

Knapp, M. and Daly, J. (2002). *Handbook of interpersonal communication*. SAGE Publications, Thousand Oaks, CA.

Kopp, S. and Wachsmuth, I. (2004). Synthesizing multimodal utterances for conversational agents. *The Journal of Computer Animation and Virtual Worlds*, 15(1):39–52.

Maxis (2000). *The Sims*. Electronic Arts.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. University of Chicago Press, Chicago, IL.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press, Chicago, IL.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7:33–35.

Musse, S. R. and Thalmann, D. (2001). Hierarchical model for real time simulation of virtual human crowds. *IEEE Transactions on Visualization and Computer Graphics*, 07(2):152–164.

Olivier, P. (2004). Gesture synthesis in a real-world eca. *Lecture Notes in Computer Science*, 3068:319–322.

Olivier, P., Jackson, D., and Wiggins, C. (2006). A real-world architecture for the synthesis of spontaneous gesture. *Proceedings of 19th annual conference on Computer Animation and Social Agents (CASA2006).*

Peters, C. and O'Sullivan, C. (2002). Synthetic vision and memory for autonomous virtual humans. *Computer Graphics Forum*, 21(4):743752.

Peters, C., Pelachaud, C., Bevacqua, E., Ochs, M., Chafai, N. E., and Mancini, M. (2006). Social capabilities for autonomous virtual characters. *GAME 2006*, pages 37–48.

Reeves, B. and Nass, C. (1996). *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, New York.

Reeves, W. T. (1983). Particle systems - a technique for modeling a class of fuzzy objects. *ACM Transactions on Graphics*, 2(2):359–376.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *SIGGRAPH Comput. Graph.*, 21(4):25–34.

Rockstar (2002). *Grand Theft Auto 3*. Take 2, Edinburgh, Scotland.

Rymill, S. J. and Dodgson, N. A. (2005). Psychologically-based vision and attention for the simulation of human behaviour. In *GRAPHITE '05: Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 229–236, New York, NY, USA. ACM Press.

Shao, W. and Terzopoulos, D. (2006). Populating reconstructed archaeological sites with autonomous virtual humans. *Proc. 6th International Conference on Intelligent Virtual Agents (IVA 06)*, 4133:420–433.

Valve (2004). *Half-Life 2*. Take 2, Bellevue, Washington.

Weisz, J. and Adam, G. (1993). Hemispheric preference and lateral eye movements evoked by bilateral visual stimuli. *Neuropsychologia*, 31(12):1299–1306.

Wilbur, M. and Roberts-Wilbur, J. (1985). Lateral eye-movement responses to visual stimuli. *Perceptual and motor skills*, 61:167–177.

Wolff, M. (1973). *Notes On The Behaviour Of Pedestrians, In: Peoples In Places: The Sociology Of The Familiar*, pages 35–48. Praeger, New York.